



Malpass, A., Dowrick, C., Gilbody, S., Robinson, J., Wiles, N., Duffy, L., & Lewis, G. (2016). Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. *British Journal of General Practice*, 66(643), e78-84.
<https://doi.org/10.3399/bjgp16X683473>

Early version, also known as pre-print

License (if available):
Other

Link to published version (if available):
[10.3399/bjgp16X683473](https://doi.org/10.3399/bjgp16X683473)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via the British Journal of General Practice at <https://doi.org/10.3399/bjgp16X683473> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Cover Page

Meaningful symptoms of low mood: problems in answer mapping and comprehension on the PHQ-9

Corresponding Author:

Dr Alice Malpass¹

Centre for Academic Primary Care

School of Social and Community Medicine

University of Bristol. BS8 2PS

Tel: 0117 9287280

Email: A.Malpass@bristol.ac.uk

Proposed Co-authors: in line with PANDA publication agreement

Chris Dowrick, Simon Gilbody, Jude Robinson, Nicola Wiles, Larisa Duffy, Glyn Lewis.

Abstract

Background: Self-administered questionnaires, such as the PHQ-9, are regularly used in clinical practice to guide prescribing or measure recovery and response to treatment. There are concerns that patients are not interpreting the questionnaire items in the same way. Cognitive interviewing is a research technique that can identify what psychologists call interpretative measurement error (IME). IME is distinct from traditional components of measurement error, such as not reading the question as worded or recording answers inaccurately.

Aims: to use cognitive interviewing techniques to explore patterns in answer mapping and comprehension of the PHQ-9 to ascertain whether the measure captures meaningful symptoms of low mood.

Design: a qualitative study using cognitive interviewing techniques and card sorting.

Methods: We recruited 18 participants at the point of entry to a longitudinal primary care depression cohort study PANDA (The indications for Prescribing ANtiDepressants that will leAd to a clinical benefit). Participants were interviewed 2, 4 and 6 weeks after their baseline visit. Cognitive interviews were digitally recorded. Analysis used the digital audio file (as opposed to verbatim transcripts) as it retained important features needed for analyses

Results: Cognitive interviewing revealed that items on the PHQ-9 are interpreted in a range of ways, that patients often cannot 'fit' their experience into the response options and so often feel the questionnaire is 'mis-representing' their experience of meaningful symptoms of low mood.

Conclusions: The PHQ-9 may be missing the presence and/or intensity of certain symptoms that are meaningful to patients. Clinicians should adopt caution when using it.

How this fits in

Whilst there are a handful of studies which have used cognitive interviewing with the Beck Depression Inventory, as far as we are aware, this is the first study to use cognitive interviewing techniques to explore answer mapping and comprehension of the PHQ-9. Research has already shown that clinicians have uncertainty about the PHQ-9's validity and utility in the management and diagnosis of depression within primary care. This study provides the first empirical evidence that the PHQ-9 may be missing the presence and/or intensity of certain symptoms that are meaningful to patients. As a result clinicians and researchers may want to continue to adopt caution when using and interpreting questionnaire scores with their patients.

Introduction

The problem:

Clinicians are encouraged to use mood questionnaires in routine primary care in a range of health settings. In the US encouragement to use mood questionnaires comes from the U.S Preventive Services Task Force (USPSTF)¹ and the Agency for Healthcare Research Quality (AHRQ)². In the UK, the Quality and Outcomes Framework (QoF), has encouraged clinician use of brief self-administered questionnaires such as the Patient Health Questionnaire (PHQ-9)³ (see Table 1). Many GPs do not think the brief severity questionnaires are valid pointers to determine treatment choices⁴ and antidepressant prescribing decisions are not based solely on reaching a threshold on the questionnaire⁵. The latest National Institute for Health and Care Excellence (NICE) guideline on depression discourages the sole use of questionnaires to guide prescription⁶. Self-report mental health questionnaires are also increasingly a focus of research⁷.

The qualitative findings presented here are part of a larger study, PANDA (the indications for Prescribing ANtiDepressants that will leAd to a clinical benefit). PANDA is a longitudinal cohort study of people with depression identified in primary care, investigating the clinically important difference on commonly used self-administered questionnaires for depressive symptoms. The PANDA study uses the “global rating of change” question⁸ approach to estimate a minimal clinically important difference. This approach takes into account the individuals’ own judgement about whether an improvement has occurred and this can then be compared with the change of scores on questionnaires-such as the PHQ-9.

From a cognitive psychology perspective⁹ comparing a global rating of change question with changes in scores on a questionnaire maybe problematic because whilst self-report measures are validated using standard quantitative approaches, they are not validated for what social theorists call ‘interpretative measurement’ error:

“the goal of standardisation is that each respondent be exposed to the same question experience so that any differences in the answers can be correctly interpreted as reflecting differences between respondents rather than differences in the (*interpretative and meaning - making*) process that produced the answer”¹⁰ (*italics added*).

Interpretative differences may be enhanced in patients with depressive symptoms. For example, patients may struggle more with memory retrieval of relevant information, inhibiting the recall of symptoms over a two week period, affecting how they map responses to the options available. Patients may comprehend the same questionnaire item in different ways because of sensitivity towards social desirability, for example not disclosing suicidal ideation¹¹. These are distinct from traditional components of measurement error, such as not reading the question as worded or recording answers inaccurately¹².

Whereas cognitive psychology is usually interested in process (comprehension and answer mapping), in this study we were also interested in the content and meaning responses had for patients. The main aim was to explore differences between the way patients comprehend and map their answer to the options on the questionnaire. A related aim was to see whether patients shift over time in how they comprehend items on the questionnaire or find them problematic to answer, perhaps in relation to their own changing symptoms.

[Insert table 1 here]

Methods

Study design

This was a longitudinal qualitative study design using cognitive interviewing techniques.

Cognitive Interviewing

Cognitive interviewing is a method to ensure respondents understand questionnaire items in a consistent way, feel able and willing to provide answers that represent their experience. Unlike most other measures, the PHQ-9 was developed and refined for use with medical patients, not psychiatric patients or community residents. This is important because the criterion validity had to be established in patients with high rates of nonspecific physical symptoms that may confound the diagnosis of Major Depressive Disorder¹³. This context helps us evaluate the sort of information its authors intended to generate. Cognitive interviewing is “used to evaluate the quality of response or help determine whether the question is generating the sort of information that its author intended”¹², including the respondents interpretation of the question and particular terms, their comfort level with answering, any meditating factors that may influence their responses (e.g. faith in God, sense of shame) and their own sense of

confidence in the accuracy or meaningfulness of their answer (i.e. does the box they tick really represent what they feel is the 'truth').

Sampling

We recruited participants who had completed baseline data for the PANDA study. Recruitment took place in Bristol. We adopted a purposive sampling strategy. This ensured ethnicity, sex and socio-demographic differences (using GP practice as a proxy for social-demographic of participants) were represented as much as possible. We selected patients with a range of Clinical Interview Schedule Revised (CIS-R) scores to represent mild, moderate and severe ICD-10 diagnosis of depression (table 2). We decided in advance to approach 20 participants as we aimed to conduct 3 cognitive interviews with each participant resulting in 60 interviews for analysis. This size of data set is large for qualitative research¹⁴. During analysis we felt we had reached a saturation of themes with 18 participants and 48 completed interviews and so did not need to continue recruitment until we reached the target of 20.

Data collection

Participants recruited to the PANDA study gave initial consent to be contacted about the qualitative study. At the first cognitive interview, participants gave full written informed consent to take part in the qualitative study. Participants were interviewed three times, 2, 4 and 6 weeks after their baseline appointment. The lead author conducted all the interviews which lasted between 50 and 180 minutes. Interviews used a protocol guide (summarised below) and were digitally recorded.

The stages of a cognitive interview: the protocol guide

Patients were invited to complete the global rating of change question and the PHQ-9 whilst thinking aloud what was going through their mind as they read the questions and pondered their answer. The lead author used non-directive, open verbal probing during this process, such as 'tell me a bit more about what you are thinking?' Observation probes were used alongside non-directive probing, such as 'you're hesitating, can you tell me why?' This was followed up with more targeted probes about the response process, for example by asking 'what does that term mean to you?'.

Card sorting is an integral part of the cognitive interviewing approach to determine how individuals organise concepts (in our case meaningful symptoms)¹². Participants were given a pack of ‘symptom’ cards, each card having one symptom from the PHQ-9 and asked to rank their symptoms on a scale of 1-10 (where 10 represents the most meaningful symptom in terms of impact or intensity and 1 represents the least meaningful). This prompted a narrative of meaningful symptoms which was digitally recorded. Blank cards were also available for participants to write symptoms that were important to them, which could be placed on the scale.

Data analysis

Analysis used the digital audio file (as opposed to verbatim transcripts) as the audio file retains important verbal features needed to contextualise analysis of ‘answer mapping’ and ‘comprehension’, such as hesitations, sighing. An excel grid was created for analysis with 18 column headings, with each column heading denoting ‘comprehension’ and ‘answer mapping’ for each item on the PHQ-9. Participants were listed in rows. Each participant had 3 rows completed, each row representing a different time point at week 2, 4 or 6. Additional columns summarised data from the card sort exercise and the global rating of change question. This approach to analysis has similarities to that used in framework analysis¹⁵.

Results

Participants

20 participants were approached, 2 participants did not respond to initial contact, 18 participants were recruited. Out of 18 participants, 14 completed all 3 interviews, two participants completed 2 interviews and two participants completed only one interview. In total, 48 cognitive interviews were completed. The age range, CIS-R scores and GP practice (as a proxy for social-demographic) of participants were evenly distributed (table 2).

[Insert table 2 here]

Overview of findings

The findings explore themes in answer mapping and comprehension using verbatim text from cognitive interviews as illustrations of an issue that, in most cases, affected participants across the sample. Each

verbatim quote is tagged with a numerical identifier, the respondents occupation and whether the data comes from the first, second or third interview. Where appropriate we refer to the card-sorting data to show under-reporting on the PHQ-9 of a symptoms intensity or impact for the participant. The card sorting exercise also invited participants to write down their own unique meaningful symptoms on blank cards. Not all patients filled them in. Those that did listed either perceptual symptoms: improvements in vitality in vision (where things look brighter and more vibrant), depersonalisation (where experience slips out of focus); feelings such as resentment, exclusion and loneliness; and somatic sensations in the body such as tremors, exhaustion, restlessness, a weight on the shoulders, pain in the body, knot in stomach, a sense of a ticking time bomb in the body and nausea. All these symptoms formed a meaningful and/or intense part of their changing low mood symptoms but were not represented on the PHQ-9. No comprehension or answer mapping issues emerged from the global rating of change question.

Problems in answer mapping and comprehension on the PHQ-9

Off-setting frequency with intensity

Participants translated the options on frequency into their own meaningful measure of intensity. For example, 'several days' was used to represent low-level intensity rather than the actual number of days a certain symptom had arisen:

"I feel sad and down sometimes, more than the average person, when I think about things I feel down every day, *if I put it nearly every day it would make it look much more severe than it really is*, because I'm not really sure, *I'd put several days because it's not committing me*, it is every day but only small parts of the day...especially now I can see more outside of the box, I can stop dwelling on the things that make me low..." (202, GP 3rd interview, italics added)

The same participant wrestled with representing intensity versus frequency of a symptom at more than one interview:

"When it's been there (feeling down, depressed and hopeless) it's been intense but it's not been as much as more than half the days, it's been intense, but it's not lasted all day... short lived but more intense" (202, GP, 2nd interview)

Similarly, another participant does not answer item 6 (feeling bad about yourself) on the basis of frequency but on the basis of the intensity and impact of her negative thoughts:

"I'm doing quite well at the moment, I'm going to put 'not at all' although there have been episodes of sitting in the car thinking 'oh god what a waste of a life, house is a mess, garden is a

mess, going to be evicted because you can't pay the rent', ruminating thoughts have been transitory, they've not settled in on me, I haven't spent that much time really thinking about myself, that nasty churning over" (181, not working, 2nd interview)

Double or triple barrelled questions

There are several triple or double barrelled questions that caused difficulty. Item 9 (Suicidal ideation) asks if patients have been bothered with "thoughts that you would be better off dead, or of hurting yourself in some way". Patients distinguished these two parts of the question as referring to very different things which made it difficult for them to answer:

"[they are] different thoughts altogether, [I'm] definitely not suicidal, just questioning God, why do you keep me alive when there is nothing here for me, suicide is self harm, but I'm asking God, 'why can I not just wake up in the morning, go in my sleep'... suicidal thoughts at Christmas were completely different feelings, feel as though you not attached to anything, you can drive a car but you don't feel like you, not hooked up to the car, driving it but not part of it, the body felt different. [example of a suicidal thought]: thinking of driving to the severn bridge and jumping off of it...don't make plans, it's just spontaneous...Thoughts that you would feel better off dead. That doesn't mean self harm does it? Does that mean suicidal thoughts? It could do, or it could be just wishing you're not here -if so I would put several days, then if it was suicidal thoughts I would put 'not at all'. If I interpret that as non-suicidal, I'll put several days" (162, volunteer at hospital, 1st interview)

Item 6 (feeling bad about yourself-or that you're a failure or have let your family down) also caused problems as participants felt they had experienced different aspects of 'feeling bad' in different frequencies and intensities:

"I do have the bad feelings about myself and those are really intense, I try to minimise the impact on family but I don't know if I always succeed...certainly the bad feeling about myself has been intense, do I have it every day? certainly the bad feeling about myself every day, it's hard because there are three aspects to that...so if was just feeling bad about myself it would be nearly every day, or that you are a failure, more than half the days, or that you've let your family down, probably several days...feeling bad about myself is a constant and the other feelings are a consequence, I'll tick every day because I can say that" (172, working mother, 1st interview)

Similarly another participant could respond to each part of item 6 with different responses:

“That’s 3 different things, if I was answering them separately, feeling down-several days, depressed-more than half the days, hopeless-not at all, I’m not hopeless because I know I can do things, that’s 3 different things... I ‘d leave that one blank...if I cross the hopeless out, I can answer it.” (188, artist, 2nd interview,)

The use of ‘or’s was confusing, leading participants to wonder, “Should I answer it if just one applies to me?” (185), or wanting to cross out the section that doesn’t apply. For example, item 4 (poor appetite or overeating):

“Poor appetite or overeating-it’s confusing because it’s got both, so I want to cross out overeating, it hasn’t affected me, only when I’m depressed, so what do I put?” (182, not working, 1st interview)

Lack of identification with examples given

Item 7 (concentration) caused comprehension problems because of the specificity of examples (intended to illustrate ‘everyday’ concentration problems) “such as watching television or reading the newspaper”. Participants often read this literally:

“That gets me as it assumes one would normally (watch TV)...I don’t normally do those things...I’d have to be a bit theoretical because I’ve not watched the television or read the newspaper...” (202, GP, third interview)

Similarly, other participants also ticked ‘not at all’ for this item because they do not read newspapers, though they described having trouble concentrating during the card sort exercise.

Comprehension issues-use of language

One participant who was never able to sleep for longer than a few hours each day, found item 3 (trouble falling or staying asleep or sleeping too much) difficult to understand and mis-represented her experience:

“I’m not getting enough sleep, so not really...‘not at all’ innit?’ Not at all’, means not sleeping as much as I am... I would like to sleep longer but I can’t, I just automatically wake up’, (Researcher probes her comprehension of the item)...i don’t have trouble falling asleep, but I wake up and

that's it I don't go back...so it would be nearly every day" (194, cleans trains overnight, 2nd interview).

Differences over time in comprehension and answer mapping

Our findings did not show that patients shift over time in how they comprehend items on the PHQ-9 or find them problematic to answer in relation to their own changing symptoms. On the contrary, the same comprehension and answer mapping problems were expressed at more than one time point by the same participants, for example, double or triple barrelled questions remained problematic overtime. However, there was a mis-match between participant perceptions of completing the questionnaire over time in relation to their symptoms. Some participants felt they had completed the questionnaire exactly the same each week because they perceived that their symptoms had not changed, but in practice their responses on the PHQ-9 had changed.

Discussion

A wide range of comprehension and answer mapping difficulties were found on the PHQ-9, which persisted over time. For example, 'language design' issues through the use of double or triple barrelled questions were problematic for those who felt they could respond differently to each part of the question. Timescale options were problematic, for example a day being experienced as so variable; or participants expressing a tension between frequency and intensity of symptoms, again making it problematic to map a meaningful answer.

As far as we are aware, this is the first study to use cognitive interviewing techniques to explore answer mapping and comprehension of the PHQ-9. The findings demonstrate the value of asking participants what meaning each item on the questionnaire had for them and their reasons for responding to each item as they did.

Limitations:

Our study has several limitations. Cognitive interviewing as a methodological approach cannot indicate the size or extent of a problem with particular items, nor can it guarantee that all problems have been captured, especially as research suggests there is a positive relation between sample size and problem detection ¹⁶.

Using cognitive interviewing techniques in a longitudinal study design may have led to participants becoming 'schooled' in the questionnaire. The use of 'non-directive' and observational probes' during questionnaire completion may have influenced how respondents continued to map their answers. However our findings showed the same issues in comprehension and answer mapping came up at each time point, suggesting participants did not adjust their answers in response to being more familiar with the questionnaire or in response to the interaction of the cognitive interview probes.

Approaches to analysis of cognitive interview data are still being developed and debated¹². The coded analysis for this study was systematic and drew upon the theoretical framework underpinning cognitive interviewing by framing analysis under 'comprehension' and 'answer mapping'⁹. Analysis was not double coded which is a limitation of the study. .

Comparison with Existing Literature

The problems identified in our study in relation to suicidal ideation items have been reported elsewhere. For example, a comparison of interview data with PHQ-9 responses found patients under-reported suicidal ideation and the measure failed to pick up increases in intensity of suicidal thought (that may be less frequent)¹¹. Our findings help explain why this under-reporting is occurring-because of the multiple ways 'thoughts of self harm' and 'being better off dead' are interpreted as statements.

Another way to view our findings is through the terms adopted by a study^{17,18} interested in the 'discursive fit' between what items demand from informants and what informants decide to do with such a demand. The research discusses three strategies informants adopt to cope with problematic items on the Beck Depression Inventory (BDI). They *reformulate* items-answering different questions from those posed by the questionnaire. They *recontextualise* items-drawing on contexts which rendered the item nonsensical. Or they *contest* the assumptions underlying the scale, rejecting it altogether. In the findings we report we can see all three strategies at work. For example, item 7 (concentration) was contested by a participant who rejected it as irrelevant because her experience did not match the examples given. Participants also contested the meaningfulness of questionnaire items repeatedly if they were double or triple barrelled questions (items 4, 6 and 9). Participants reformulated the options in frequency (not at all, several days, half the days, more than half the days) into their own personalised scale of intensity.

Implications for research and practice:

Our findings suggest that the wording on the PHQ-9 could be improved so that patients and clinicians can more usefully distinguish between frequency and severity of symptoms. Research shows that patients who get better while undergoing treatment score better on the PHQ-9, indicating them as a reliable measure of patients' condition and recovery³. How do we reconcile psychometric credibility based upon quantitative measures of reliability and validity with qualitative analysis, such as ours, that raises questions about its utility as a measure to represent symptoms that are meaningful to patients?

One plausible explanation is that patients in clinical settings (or research settings) are not encouraged to challenge or comment upon the questionnaire, as participants are in cognitive interview studies, and instead routinely engage in "trying to give the 'right' answer", knowing what is at stake (p.528)¹⁷ and so adopt a "fake-good profile"(p.497)¹⁹. The following commentary on the BDI may equally apply to the PHQ-9:

"The BDI works within the parameters of the dominant discourse of psychiatry and clinical psychology and so it successfully measures something, because it corresponds with the rules of what constitutes such measurement. And while it might identify (Major) Depressive Episode (ICD F32-33 or DSM 296.2-3) *it is unlikely to pin down the individual experience of low mood, sadness or what we call 'depression'*" (p.529, italics added)¹⁷.

Patients complete the PHQ-9 in socially situated and power laden contexts. Researchers stress the importance of qualitative methods in the ongoing evaluation of instruments, to inform quantitative psychometric evaluations and the appropriate use of instruments in clinical practice¹⁹. Our findings are of relevance to ongoing clinical practice because they suggest, as clinicians have suspected for some time, that screening measures are limited when compared to practical wisdom and clinical judgment⁵. Clinicians have expressed uncertainty about the PHQ-9's validity and utility; and in the management and diagnosis of depression within primary care have a strong preference for clinical judgment over scores on severity measures⁵.

Conclusion

The study reminds us of the importance of following routine advice in questionnaire design, such as avoiding double or triple barrelled questions. What this study demonstrates most clearly is that the "quest for meaning and the quest for measurement are incommensurable research objectives" (p.411)¹⁹. In light of the numerous ways the PHQ-9 may be missing the presence and/or intensity of certain symptoms that are meaningful to patients, clinicians should continue to adopt caution when

using and interpreting questionnaire scores. The study raises the question that longer assessments may be better in providing opportunities for distinguishing frequency and severity, for example, as the CISR does.

Acknowledgement and Disclaimer

This is a summary of independent research funded by the National Institute for Health Research (NIHR)'s Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0610-10048). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- 1 US Preventive Services Task Force. Screening for depression in adults: US preventive services task force recommendation statement. *Ann Intern Med*, 2009. 151(11): 784-92.
- 2 Panel, D. G. (1993). Depression in Primary Care: Volume 2. Treatment of Major Depression. Clinical Practice Guideline, Number 5.(AHCPR Publication No. 93-0551). Rockville, MD: US Department of Health and Human Services. *Public Health Service, Agency for Health Care Policy and Research*.
- 3 Kroenke K, Spitzer RL, Williams JB; The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-13
- 4 Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ*. 2009. 338:b750.
- 5 Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, et al. Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ*. 2009. 338:b663.
- 6 Excellence NifHaC. Depression (updated edition). London: The British Psychological Society and The Royal College of Psychiatrists, 2009.

- 7 Roy, T., & Lloyd, C. E. . Cultural Applicability of Screening Tools for Measuring Symptoms of Depression. In *Screening for Depression and Other Psychological Problems in Diabetes* 2013. (pp. 67-86). Springer London.
- 8 Harmer CJ, Cowen PJ, Goodwin GM. Efficacy markers in depression. *J Psychopharmacol*. 2011. 25(9):1148-58.
- 9 Tourangeau, R., Rips,L., Rasinski,K. The psychology of survey response. 2000. Cambridge. Cambridge University Press.
- 10 Floyd J. Fowler, & Thomas W. Mangione (Eds.). *Standardized survey interviewing: Minimizing interviewer-related error*.1990.. Sage.
- 11 Malpass, A.,Shaw,A., Kessler,D, sharp,D. Concordance between PHQ-9 scores and patients' experiences of depression. *BJGP*. 2010. e231-e238.
- 12 Willis,G.B. 2005. Cognitive Interviewing: a tool for improving questionnaire design. Sage, London
- 13 Bombardier, C. H., Richards, J. S., Krause, J. S., Tulskey, D., & Tate, D. G. . Symptoms of major depression in people with spinal cord injury: implications for screening. *Archives of physical medicine and rehabilitation*, 2004. 85(11), 1749-1756.
- 14 Guest, G., Bunce, A., & Johnson, L.. How many interviews are enough? An experiment with data saturation and variability. *Field methods*, 2006.18(1), 59-82.
- 15 Ritchie, J., & Spencer, L. . Qualitative data analysis for applied policy research. *The qualitative researcher's companion*.2002. 305-329.
- 16 Blair, J., Conrad, F., Ackermann, A. C., & Claxton, G. (2006, May). The effect of sample size on cognitive interview findings. In *Proceedings of the American Statistical Association*.
- 17 Galasinski, D. Constructions of the self in interaction with the Beck Depression Inventory, *Health*, 2008. 12 (4) 515-533.
- 18 Galasinski,D., Kozłowska,O. Questionnaire and lived experience: strategies of coping with the quantitative frame. *Qualitative Inquiry*. 2010. 16 (4), 271-284.
- 19 Barroso,J., Sandelowski,M. In the field with the Beck Depression Inventory. *Qualitative Health research*. 2001.11(4), 491-504.

